

#2 OIPE

RAW SEQUENCE LISTING

DATE: 12/07/2001

PATENT APPLICATION: US/09/997,664

TIME: 13:03:07

Input Set : A:\BC1018 US CIP Seq Listing.txt

Output Set: N:\CRF3\12072001\I997664.raw

ENTERED

3 <110> APPLICANT: Ben-Bassat, Arie
 4 Cattermole, Monica
 5 Gatenby, Anthony A.
 6 Gibson, Katherine J.
 7 Ramos-Gonzalez, Isabel
 8 Ramos, Juan
 9 Sariaslani, Sima
 11 <120> TITLE OF INVENTION: Method for the Production of p-Hydroxybenzoate in Species of
 12 Pseudomonas and Agrobacterium
 14 <130> FILE REFERENCE: BC1018 US CIP
 C--> 16 <140> CURRENT APPLICATION NUMBER: US/09/997,664
 C--> 17 <141> CURRENT FILING DATE: 2001-11-28
 19 <150> PRIOR APPLICATION NUMBER: 09/585,174
 20 <151> PRIOR FILING DATE: 2000-06-01
 22 <160> NUMBER OF SEQ ID NOS: 112
 24 <170> SOFTWARE: Microsoft Office 97
 26 <210> SEQ ID NO: 1
 27 <211> LENGTH: 6491
 28 <212> TYPE: DNA
 29 <213> ORGANISM: Pseudomonas mendocina KR-1
 31 <400> SEQUENCE: 1
 32 tcaactccccc ttgagccggt agctgatctg cgcgcgactc atgcccaaca tctgcgcgcg 60
 33 cgcggtgagg ttgccgcggg tgcgttccag ggcaaggtgc accagcgctc gctcgatctc 120
 34 cttcagtgat gtgcctagta cccggtcgcg cccggcgagg aaggcctgca ggttgccag 180
 35 cccagctca accggttcat gtcctcgac aaccacctca gccgcgctt gcggttcgcc 240
 36 gccgacggca tccagacggc cttcggcggt caggccgatg ccgctggagc gaagtggctc 300
 37 gccggctttt gccaggtgca ccaggtcgat cagctcgcca ctgcctgcgg cgatcacgcc 360
 38 gcgctcgatc aggttctgca gtcacggat attgcgggg aagcggtagg tcagcagcgc 420
 39 gttgaccagc cgcgtgctga aaccagggg tttgagccca tggcgcgcac tgaacttgcg 480
 40 caggaagtag ctcacagga gcgggatgtc ctcacggcgc tcgcgcaggg gcggcagatg 540
 41 gatggggaac acgttcagcc ggtacagcag gtcctcgcg aagcgcggcg cctcgacctc 600
 42 tcggcgcagg tccagattgg tggcggcgat caccctcaca tccaccggga tcgcccaggt 660
 43 accacctacc cgctcgatct cgcctcctg cagcacccgc aggatcttgc tctgggcgct 720
 44 gaggtcagg gtggcgatct cgtcgaggaa cagggtgccg cccttggccc gctcgaagcg 780
 45 ccccgggcgg gaacggtcgg cgcgggtgaa ggcaccgcgc tccacgccga acagttcggc 840
 46 ttccagcaga gtttcggca acgcccgcga gttgagcgcc accaacggcg tttggcggcg 900
 47 cgggctggcc tgggtgcaggg tgcgcgcgaa gagctccttg cccacccccg attcaccggt 960
 48 cagcagtagc gtggcctggg tcgacgcaac gcggtagagc tgctggctgg cggcgacaaa 1020
 49 ggcgggcgaa atgccacca tggcctggtc ctcgggcggc tcacccagat cggccatttc 1080
 50 cgtctcgtct gccgagccgt aggtgtccg gctgaggaag tcgctggcat ccagggtggc 1140
 51 caggtcgggt tcgatgtct cccactgctc cgcggcttg ccgacgatgc ggcacgccga 1200
 52 atggccccat cagcggcatt cctgctcgcg gaaccacc aggcgcccc gaggaggga 1260
 53 ggtgtgacgg ctggcgtagc ccacttccat ccagcaggcc ggttcgctgc ccagcccgtg 1320
 54 gctggcgatg tgcctgctcg cttccaggga gttgtgccag aagaattcgg aatagaaatg 1380
 55 cccgatgctg gagtcgatgt cgaagcgcac cacttccacg ttcaccatgc cctccagcat 1440
 56 gtgcaggcgc gggcctgcgc tgtagaggct ggcgtggctg ccctcggggc actgcgcgct 1500
 57 gacctgagcg gcattcctcg ttccggcctg ccagccaatg cgggtcagaa ggccacgggc 1560

RAW SEQUENCE LISTING

DATE: 12/07/2001

PATENT APPLICATION: US/09/997,664

TIME: 13:03:07

Input Set : A:\BC1018 US CIP Seq Listing.txt

Output Set: N:\CRF3\12072001\I997664.raw

```

58 cttgtcgagg ccgagggtt ccaccaactc gcgacggatg gcgccgaagg cggccccctg 1620
59 cagcagcatc atgcgctggc cgcagagcca gatattgcca tcctggggcg cgaaggcgac 1680
60 ggtctccgcc agttgctcgg ccgagggcag cccgctgctg ccgaactggt tggcctggtc 1740
61 gacgatcagg ctcttctgcc ggccgagcaa ctgcttgagg aattcatccc ccatgctgcg 1800
62 gccgggattg ctcgagggtt tgcgagtcac ggtcatgggg cgggaggtag gaacaatgtt 1860
63 attcagtatg cccgtgtgaa atggccggtc aattggccct tgccatcacc caataatcgc 1920
64 ccaacctctt gcagaccact ccggagaagt ttctgcgccc cggagacttc tctgaagaaa 1980
65 aatcggcgcc aacctcccg caagcccccc atgcgtccgc tccgcattcc ccaaaaaaac 2040
66 gtaaccaatt gttttacaaa taaaaaatag aagaaagaag gattggcacg gtagttgtta 2100
67 aaggacaggg gcggtgcacc aagacaataa caacacaggt aacgacccta tgaaccgctt 2160
68 cccatcgcca atccattcog cttgccacc cgcaccacgg cttcgttggt gaccctcaac 2220
69 cgtacctcca caggaacggc gcccgcgctt cttgcctgac gtatcgccac gcgcccgtgt 2280
70 aaccaccggc tcgcgcgccac tggcagcctt ccgcgcaaac aagagagaac ccatggacac 2340
71 caccgcctt gcctaccaga acctcgagct ccaacctctc gccgggcaat ggcgcgcgg 2400
72 cagtagcggg cgcgcgttgg aggtcttcga cccctacaac gacgagctgc tattgcgcac 2460
73 cgccctggcc agccgcgaag acctcgacgc agcctaccgc aaggcccgcg acagccagcg 2520
74 ggagtggggc accacggcgc cggccgagcg cgcgcgggtg ctgctggaag cgggtgaagat 2580
75 cttcgatgag cgcgcgaggg agattatcga ctggatcacc cgcgagtcgg gcagcaccgg 2640
76 catcaaggcg cagatcgaat ggggcgcgcg ccgcgccatc accctggagt cggccagcct 2700
77 gccgaatgcg gtgcacgggc gcatactgc ctccaacatc tccggcaagg agagcccgct 2760
78 gtaccgcgcg cccctggggc tgatcggcgt gatcagtcgg tggaaacttc ccctgcacct 2820
79 cactgcccgc tccctggccc cggccctggc cctgggcaat gccgtggtgg tcaagccggc 2880
80 cagcgacacc ccgatcccg gtggcctact gctggcgcgc atcttcgaag aagccggcct 2940
81 gccggcgggc gtgctcagcg tgggtggtgg ttccggcgcg gagattggtg acgccttcgt 3000
82 cgagcaccgg gtgcccgcgc tcaatttcct caccggctcc actcaggtgg gccgcaacat 3060
83 cggccgcacc gccagcggcg gtgagcacct caagcacgtg gcgctggaac tgggcggcaa 3120
84 cagcccgttt gtggtcttgg ccgatgccga cgtggagcag gcggtgaatg cggccgtggt 3180
85 cggcaagttc ctgcaccagg gccagatctg catggcgatc aaccgcatta tcgtcgagca 3240
86 gcctttgctg gaagatttca cccgcgcgtt cgtcgagcgc gtcaaggccc tgccctatgg 3300
87 cgacccgagc aagccgggga ccgtggtcgg tccggtgatc aacgccaggc agctggccgg 3360
88 tctgaaggag aagatcgcca ccgccaaagg cgaaggcgcc accctgctgc tgggtggcga 3420
89 gcccagggc aacgtcatgc cgcctcatgt gttcggcaac gtcaccgccg acatggaaat 3480
90 cgcccgcgaa gaaattttcg gcccgctggt gggcatccaa tccgcccgtg acgccgaaca 3540
91 cgccctggag ttggccaaca gcagcgagta cggcctgtcc agcgcggtgt tcaccgccag 3600
92 cctcgagcgc ggcgtgcagt tcgcccggcg catccacgcc ggcatagccc acgtgaacga 3660
93 catcccggtt aacgacgagc ccaacgctcc cttcggcgcg gagaagaact ctggcctcgg 3720
94 ccgcttcaac ggcgactggg ccatacgagg gttcaccacc gatcactgga tcaccctgca 3780
95 acacagcccg cggccctatc cgttctgatg ctgcgcgcat cccatcaccg agccccaata 3840
96 aaaaacggag tacgaaatgt cctcactcct caacagccga gctgtgaaac ggccactgct 3900
97 ggccagcctt gactaattt tcgcccgtgt cgcgggccag gccttcgccg acggcgacgg 3960
98 cgtctggaaa ggcggcgaga acgtctacca gaaaatctgt ggccactgcc acgaaaaaca 4020
99 ggtgggcccg gtgatcaccg gccgccagct accgcgcgag tacatcagtg ccgtggtgcg 4080
100 caacggcttc cgcgcatatg cggcctttcc ggcctcgttc atcgacgaca aggccttgca 4140
101 gcaggtcgcc gactacatct ccaagacccc tgctactgtg gccaaagcct gaggtgccgg 4200
102 cgatgaacat cgaacgtcgt accctgctca agggcatggc cctgggcggc ctggctggcg 4260
103 ccgcatggg cgccttcggc ctggcgatga ccaaggccat gctgggcggg caggcccagc 4320
104 caetgcccac cctcgtcctg gtagatggcg aggcggccgg agcggccttc ctgcgccgag 4380
105 tcggttccag cccggcggcc agcaaggccg aggtgcagcg caccgatctc ggccctggact 4440
106 tcgtcttggg cctggagaag cgcctgcgca gtggtcagca gcaacgcac atcgggtctg 4500

```

RAW SEQUENCE LISTING

DATE: 12/07/2001

PATENT APPLICATION: US/09/997,664

TIME: 13:03:07

Input Set : A:\BC1018 US CIP Seq Listing.txt

Output Set: N:\CRF3\12072001\I997664.raw

```

107 tggatgacgc cagcgccgct ctgatacctcg acctggccccg cagcagcggc gcgcgggtgc 4560
108 agtggctcgg ccagcatagc gccgcggccg gctcctcccg gcaccgtctg ctcagcgccg 4620
109 acagcgccca gggctgctcc ctctgcctgg gccagcagct ccattgcctgc ggcgggcggt 4680
110 tcagcctgag cgaacagcac cccctgggtg gccagccctt gaattctggcc ggtgcgcgcg 4740
111 gcagcggcgg ctccgcgcaa tgggcggcca gcatacgcca cgacctggcc agcctgggcg 4800
112 gcgatgacag cagtgcggcc ccacgcattg ccaaccatta cccggcgctt accggccaat 4860
113 tcgtttcgtt ctgatcctg gtttgaagga gctgacagat gaccgagcaa acccagaaca 4920
114 cctgattcc cctggcggtg aatgacgcca acctccagca agccctggcc aagttccgca 4980
115 agctgctggg cgaggacaac gtcctggtca aggacgagca actcatcccc tacaacaaga 5040
116 tcatgatcgc agtggacaac gccgaacacg cgccctccgc tgcgtgcacc gccaccactg 5100
117 tggaacaggt gcagggcggt gtgaagatct gcaacgaata cggcattccg gtgtggacca 5160
118 tctccaccgg ccgcaacttc gggtacggct cggcgccccg cggccagcgt ggccagggtga 5220
119 tcctcgacct gaagaaaatg aacaagatca tccacgtaga cccggacctg tgcaccgccc 5280
120 tgggtgaacc gggggtgacc taccagcagc tgtacgatta cctggaagag aacaacatcc 5340
121 cgctgatgct gtctttctct gcaccctcgg ccatacgccg cccgctgggc aacaccatgg 5400
122 accgtggcgt gggctacacc ccctacggcg agcacttctt catgcagtgc ggcattggaag 5460
123 tgggtgctgg caatggcgac gtctaccgca ccggcatggg cggggtgaaa ggcgacaacg 5520
124 cctggcaggt gttcaagtgg ggctacggcc cgacctgga cggcatgttc acccaggcca 5580
125 actacggcat ctgcaccaag atgggtttct ggctgatgcc caagcccccg gtgttcaagc 5640
126 ccttcgagat caagttcgag aacgagtcgg acatcagcga gatcgtcgaa ttcattccgtc 5700
127 cgctgcgcat cgcccaggtc atcccaaact ccgtggtgat cgccggtgtg ctctgggagg 5760
128 cctccacctg caataccgcg cgctcgact acaccactga gccgggcgcc actcccgaca 5820
129 ccatactgaa gcagatccag aaggacaagg aactcggcgc ctggaacgtc tatgccgtc 5880
130 tctacggcac gcaggaacag gtggacgtga actggaagat cgtcaccggc gccctggcca 5940
131 aactgggcaa gggcaggatt gtcacccagg aagaggccgg cgatacccag cccttcaagt 6000
132 accgttccca gttgatgtcc ggcgtcccca acctgcagga attcggcctg tacaactggc 6060
133 gcgggggcgg cggctccatg tggttcgccc cggtcagcca ggcccggtgc atcgagtgcg 6120
134 acaagcagca ggcgtggcc aagaagatcc tcaacaagca cggcctggac tacgtcggcg 6180
135 agttcattgt cggctggcgc gacatgcacc acgtaatcga cgtgctgtac gaccgcacca 6240
136 accccgagga aacccaacgc gcctacgcct gcttccacga gttgctggat gagttcgaga 6300
137 agcacggcta tgcggtgtac cgcgtgaaca ctgcgttcca ggagcgcgtg gcgcagaggt 6360
138 acggcacggt caagcgcagg tggaacacgc catcaagcgc gccctggacc cgaacaacat 6420
139 cctggcaccc ggcaaatccg gcatcgacct cgccaacaag ttctaaccct aagcaagacc 6480
140 ccgccgggta a 6491

```

142 <210> SEQ ID NO: 2

143 <211> LENGTH: 611

144 <212> TYPE: PRT

145 <213> ORGANISM: Pseudomonas mendocina KR-1

147 <400> SEQUENCE: 2

```

148 Met Thr Met Thr Arg Lys Pro Ser Ser Asn Pro Gly Arg Ser Met Gly
149   1           5           10           15
151 Asp Glu Phe Leu Lys Gln Leu Leu Gly Arg Gln Lys Ser Leu Ile Val
152           20           25           30
154 Asp Gln Ala Asn Gln Phe Gly Ser Ser Gly Leu Pro Ser Ala Glu Gln
155           35           40           45
157 Leu Ala Glu Thr Val Ala Phe Ala Pro Gln Asp Gly Asn Ile Trp Leu
158           50           55           60
160 Cys Gly Gln Arg Met Met Leu Leu Gln Gly Ala Ala Phe Gly Ala Ile
161  65           70           75           80

```

RAW SEQUENCE LISTING

DATE: 12/07/2001

PATENT APPLICATION: US/09/997,664

TIME: 13:03:07

Input Set : A:\BC1018 US CIP Seq Listing.txt

Output Set: N:\CRF3\12072001\I997664.raw

```

163 Arg Arg Glu Leu Val Glu Ala Leu Gly Leu Asp Lys Ala Arg Gly Leu
164      85      90      95
166 Leu Thr Arg Ile Gly Trp Gln Ala Gly Thr Arg Asp Ala Ala Gln Val
167      100     105     110
169 Ser Ala Gln Trp Pro Glu Gly Asp His Ala Ser Leu Tyr Ser Ala Gly
170      115     120     125
172 Pro Arg Leu His Met Leu Glu Gly Met Val Asn Val Glu Val Val Arg
173      130     135     140
175 Phe Asp Ile Asp Ser Ser Ile Gly His Phe Tyr Ser Glu Phe Phe Trp
176 145      150     155     160
178 His Asn Ser Leu Glu Ala Asp Glu His Ile Ala Ser Tyr Gly Leu Gly
179      165     170     175
181 Ser Glu Pro Ala Cys Trp Met Glu Val Gly Tyr Ala Ser Gly Tyr Thr
182      180     185     190
184 Ser Ser Leu Leu Gly Arg Leu Val Val Phe Arg Glu Gln Glu Cys Arg
185      195     200     205
187 Cys Met Gly His Ser Ala Cys Arg Ile Val Gly Lys Pro Ala Glu Gln
188      210     215     220
190 Trp Glu Asp Ile Asp Thr Asp Leu Ala His Leu Asp Ala Ser Asp Phe
191 225      230     235     240
193 Leu Ser Arg Ser Thr Tyr Gly Ser Ala Asp Glu Thr Glu Met Ala Asp
194      245     250     255
196 Leu Asp Glu Pro Pro Glu Asp Gln Ala Met Val Gly Ile Ser Ala Ala
197      260     265     270
199 Phe Val Ala Ala Ser Gln Gln Leu Tyr Arg Val Ala Ser Thr Gln Ala
200      275     280     285
202 Thr Val Leu Leu Thr Gly Glu Ser Gly Val Gly Lys Glu Leu Phe Ala
203      290     295     300
205 Arg Thr Leu His Gln Ala Ser Pro Arg Arg Gln Thr Pro Leu Val Ala
206 305      310     315     320
208 Leu Asn Cys Ala Ala Leu Pro Glu Thr Leu Leu Glu Ala Glu Leu Phe
209      325     330     335
211 Gly Val Glu Arg Gly Ala Phe Thr Gly Ala Asp Arg Ser Arg Pro Gly
212      340     345     350
214 Arg Phe Glu Arg Ala Lys Gly Gly Thr Leu Phe Leu Asp Glu Ile Ala
215      355     360     365
217 Thr Leu Ser Leu Ser Ala Gln Ser Lys Ile Leu Arg Val Leu Gln Glu
218      370     375     380
220 Gly Glu Ile Glu Arg Val Gly Gly Thr Ser Ala Ile Pro Val Asp Val
221 385      390     395     400
223 Arg Val Ile Ala Ala Thr Asn Leu Asp Leu Arg Arg Glu Val Glu Ala
224      405     410     415
226 Gly Arg Phe Arg Glu Asp Leu Leu Tyr Arg Leu Asn Val Phe Pro Ile
227      420     425     430
229 His Leu Pro Pro Leu Arg Glu Arg Arg Glu Asp Ile Pro Leu Leu Met
230      435     440     445
232 Ser Tyr Phe Leu Arg Lys Phe Ser Ala Arg His Gly Leu Lys Pro Leu
233      450     455     460
235 Gly Phe Ser Thr Arg Leu Val Asn Ala Leu Leu Thr Tyr Arg Phe Pro

```

RAW SEQUENCE LISTING

PATENT APPLICATION: US/09/997,664

DATE: 12/07/2001

TIME: 13:03:07

Input Set : A:\BC1018 US CIP Seq Listing.txt

Output Set: N:\CRF3\12072001\I997664.raw

```

236 465          470          475          480
238 Gly Asn Ile Arg Glu Leu Gln Asn Leu Ile Glu Arg Gly Val Ile Ala
239          485          490          495
241 Ala Gly Ser Gly Glu Leu Ile Asp Leu Val His Leu Ala Lys Ala Gly
242          500          505          510
244 Glu Pro Leu Arg Ser Ser Gly Ile Gly Leu Thr Ala Glu Gly Arg Leu
245          515          520          525
247 Asp Ala Val Gly Gly Glu Pro Gln Ala Arg Ala Glu Val Val Val Glu
248          530          535          540
250 Glu His Glu Pro Val Glu Leu Gly Leu Ala Asn Leu Gln Ala Phe Leu
251 545          550          555          560
253 Ala Gly Arg Asp Arg Val Leu Gly Thr Ser Leu Lys Glu Ile Glu Gln
254          565          570          575
256 Arg Leu Val His Leu Ala Leu Glu Arg Thr Gly Gly Asn Leu Thr Ala
257          580          585          590
259 Ala Ala Gln Met Leu Gly Met Ser Arg Ala Gln Ile Ser Tyr Arg Leu
260          595          600          605
262 Lys Gly Glu
263          610
265 <210> SEQ ID NO: 3
266 <211> LENGTH: 491
267 <212> TYPE: PRT
268 <213> ORGANISM: Pseudomonas mendocina KR-1
270 <400> SEQUENCE: 3
271 Met Asp Thr Thr Arg Pro Ala Tyr Gln Asn Leu Glu Leu Gln Pro Leu
272 1          5          10          15
274 Ala Gly Gln Trp Arg Ala Gly Ser Ser Gly Arg Pro Leu Glu Val Phe
275          20          25          30
277 Asp Pro Tyr Asn Asp Glu Leu Leu Leu Arg Ile Ala Leu Ala Ser Arg
278          35          40          45
280 Glu Asp Leu Asp Ala Ala Tyr Arg Lys Ala Arg Asp Ser Gln Arg Glu
281          50          55          60
283 Trp Ala Thr Thr Ala Pro Ala Glu Arg Ala Arg Val Leu Leu Glu Ala
284 65          70          75          80
286 Val Lys Ile Phe Asp Glu Arg Arg Glu Glu Ile Ile Asp Trp Ile Ile
287          85          90          95
289 Arg Glu Ser Gly Ser Thr Arg Ile Lys Ala Gln Ile Glu Trp Gly Ala
290          100          105          110
292 Ala Arg Ala Ile Thr Leu Glu Ser Ala Ser Leu Pro Asn Arg Val His
293          115          120          125
295 Gly Arg Ile Ile Ala Ser Asn Ile Ser Gly Lys Glu Ser Arg Val Tyr
296          130          135          140
298 Arg Ala Pro Leu Gly Val Ile Gly Val Ile Ser Pro Trp Asn Phe Pro
299 145          150          155          160
301 Leu His Leu Thr Ala Arg Ser Leu Ala Pro Ala Leu Ala Leu Gly Asn
302          165          170          175
304 Ala Val Val Val Lys Pro Ala Ser Asp Thr Pro Ile Thr Gly Gly Leu
305          180          185          190
307 Leu Leu Ala Arg Ile Phe Glu Glu Ala Gly Leu Pro Ala Gly Val Leu

```

VERIFICATION SUMMARY

PATENT APPLICATION: US/09/997,664

DATE: 12/07/2001

TIME: 13:03:08

Input Set : A:\BC1018 US CIP Seq Listing.txt

Output Set: N:\CRF3\12072001\I997664.raw

L:16 M:270 C: Current Application Number differs, Replaced Application Number

L:17 M:271 C: Current Filing Date differs, Replaced Current Filing Date